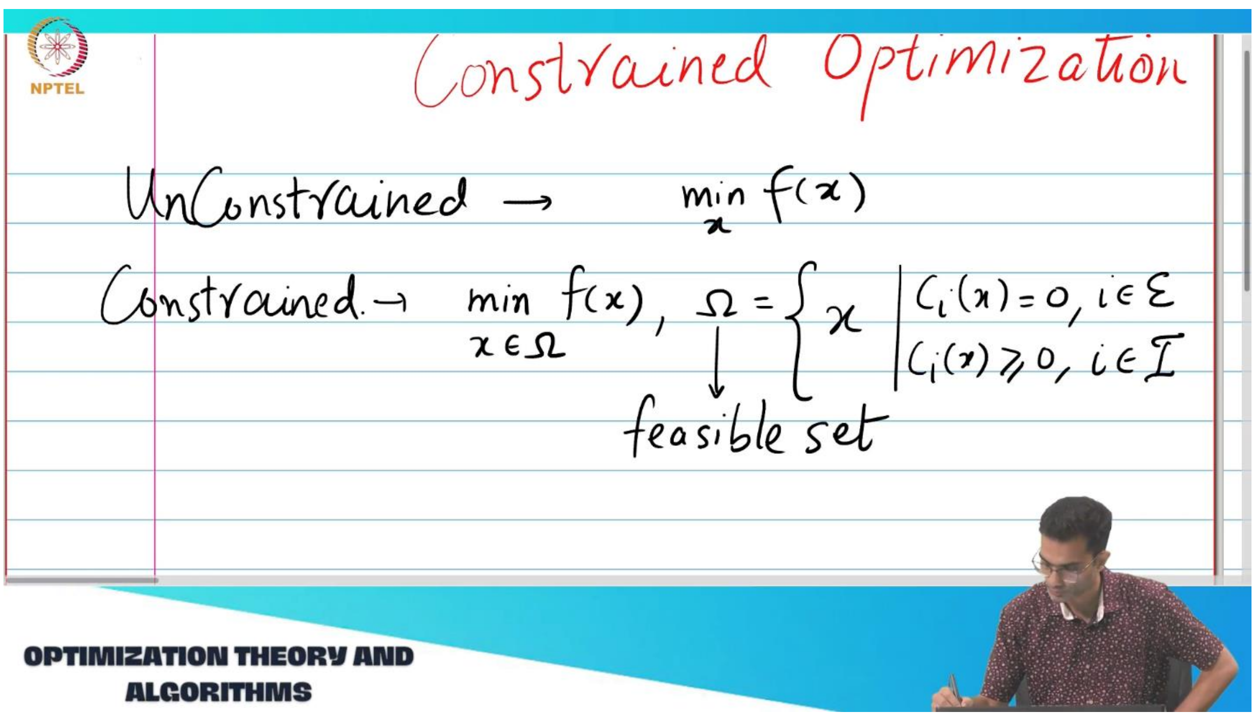
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 09
Lecture - 61

Single equality constraint

So, we are going to resume our attempt at constraint optimization from last time, ok. So, very very brief overview of what we had was, what is the terminology of constraint optimization? We had that the constrained was simply written as, right, this was our sorry unconstrained and constrained became. How did I write it? I said, I said x belongs to a feasible set Σ. And how did I define this capital Σ? We said, we wrote it in terms of equalities and inequalities. So, this is our just the way we formulated. What did we call Σ as? Feasible set. Having done this now we will start with our explanation of the simplest kind of problem to start with which was equality constraint.

One equality and the rest is constraint ok. So, as before our objective function $x_1 + x_2$ and our constraint equality was simply what? The circumference of a circle. This was our constraint. That is what I had, yeah right.

Now for this problem, it is good to draw a geometry for this. This is my $C_1(x) = 0$. What was we calculated $\nabla C_1(x)$, what was that? $2x_1, 2x_1x_2$. and geometrically this was what way? Always pointing radially outwards, right. So, take a few points like this.

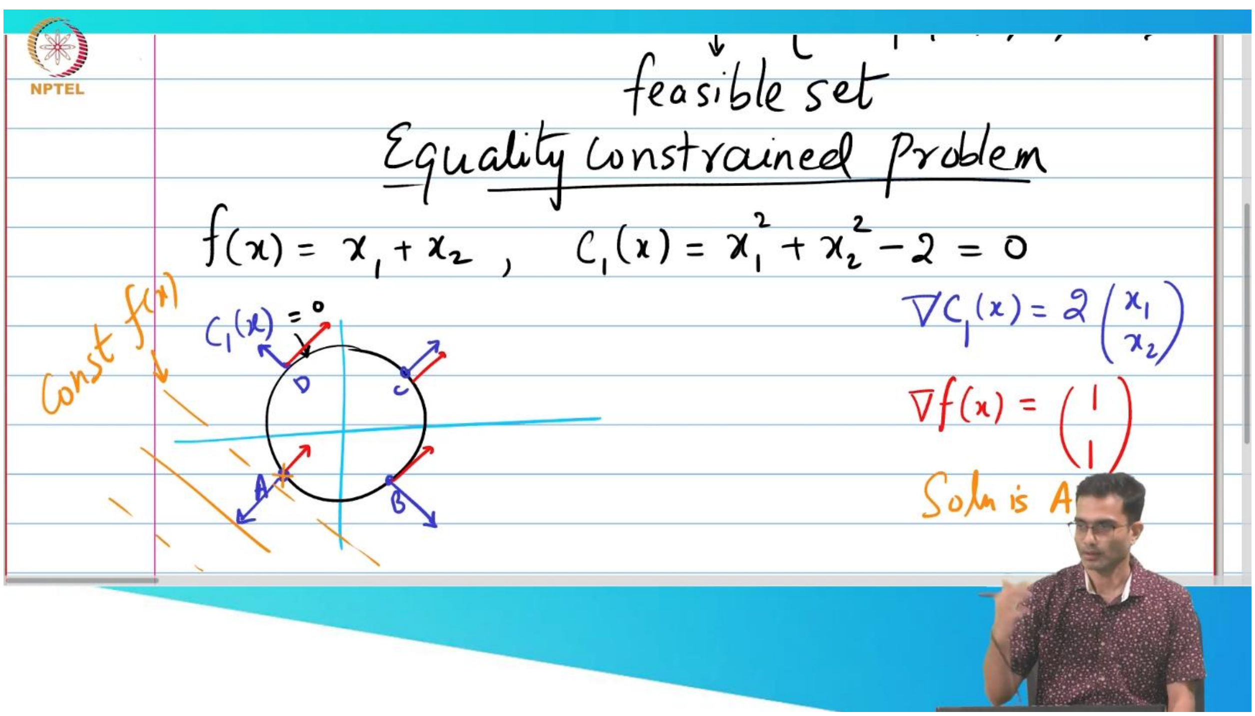


If I take 4 points over here, A, B, C, and D. So, this is the way ∇C_1 points and then I can also look at $\nabla f(x)$ and that turns out to be $\frac{df}{dx_1}, \frac{df}{dx_2}$. This is a constant vector, so it is always pointing in the same direction, ok. Then the other thing is what are the contours of constant $f(x)$. So, this is very simple to see.

What are the contours? Straight lines, parallel lines with a slope of $-45°$ you can say. So, these are the contours of constant $f(x)$, ok. And which way is $f(x)$ increasing towards the top right or bottom left? It is increasing towards the top right. So, we know already looking at this diagram that the solution is going to be because that is the place where $f(x)$ is the minimum and yet I am on the constraint right. So, without solving this problem in any formal way I know that the solution is easy.

So, this is a very simple problem I do not need any of the machinery of constraint optimization for this, but this will give me the geometric intuition to formulate how to solve this problem. So, what was the, what is the observation that we can make at the solution? So, at A if you were to try to come up with some geometric relation between ∇f and ∇C_1 you can see that they are parallel to each other or anti-parallel. So, you could simply write it as $\nabla f(x^*) = \lambda \cdot \nabla C_1(x^*)$. right. This is, is this true at points B and D? No, it is not true. But it is true at point C, ok.

And that is in a way ok because when we did our unconstrained optimization, what was our condition for stationary point in unconstrained optimization? $\nabla f(x^*) = 0$ and that was true at a maxima as well as a minima. So, nothing special over here. We are starting with the simplest. What kind of a condition would you call this? Could be a necessary condition. It is not going to be sufficient.



So, this is the analog of that. I am coming up with a necessary condition that a minima point should show. Point C, if you were to characterize it, what is point C? It is actually a maxima. At which $f(x)$ is maximized while satisfying the constraint. So, it is not that much of a surprise that ∇f is 0 at that point also.

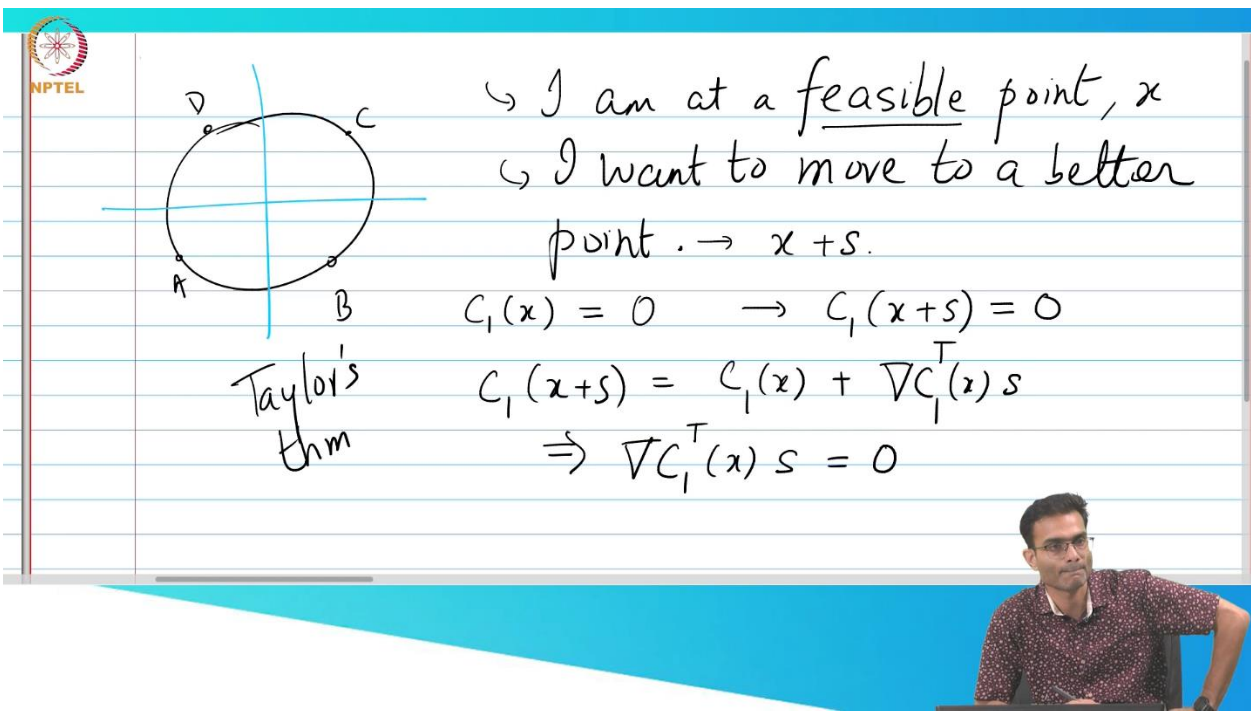
But let us try to now formalize this a little bit more. Let us draw our geometry once again. So now let me ask you a question. Let us say I am at some point and obviously I am going to assume I am not at the true. So I am at A, I have to assume that I am at a feasible point. So this I mentioned this earlier, this is a very important assumption.

We are going to assume that at every point of the algorithm I am at a feasible point. I am not somewhere out in deep space. So I am at a feasible point. In this case, it means that I am on the circle.

Okay. So $C_1(x) = 0$ is always going to be satisfied. That is a given. Okay. So now the question is I want to move to a better point.

Okay. So I want to move. Okay. So let us say I am at a feasible point x and I want to move to a point $x + s$. Obviously s is a vector of the same dimension. So $C_1(x)$, what is the value of $C_1(x)$? 0 because I am at a feasible point.

I have moved to a new point $C_1(x + s)$, what will C_1 be? Has to be also 0 because I am only going along feasible points. So, this is already telling me something that should happen. If I apply Taylor's theorem to C_1. So, Taylor's theorem. So, $C_1(x + s)$, first term will simply be $C_1(x)$, second term will be $\nabla C_1(x)^T \cdot s$.



This is our simple, right. And so this tells me what? There is no option, but $\nabla C_1(x)^T \cdot s = 0$. $\nabla C_1(x)^T = 0$. That is how it looks algebraically. Geometrically what does this mean? That means this direction s in which I am going must be orthogonal to ∇C_1.

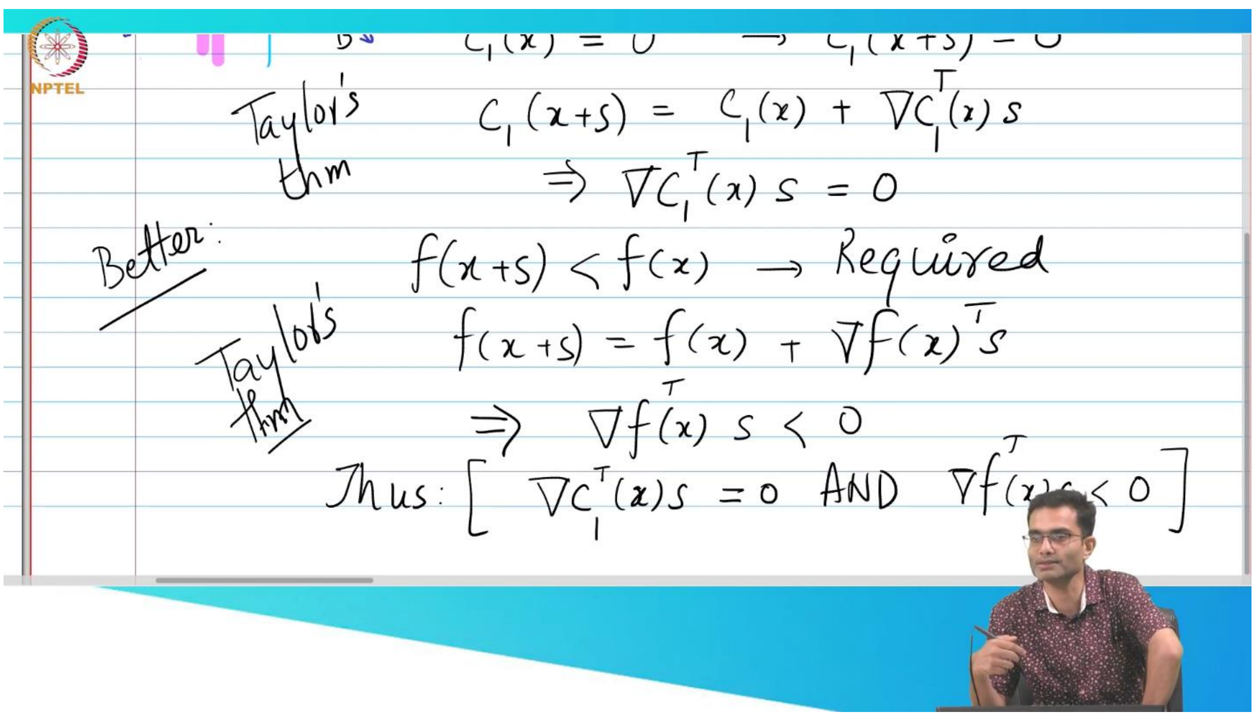
And ∇C_1 as you saw in the previous slide was always pointing radially outwards. So, I know this. So, let us now pick a point. Let us say I am at this point. This is my current point x. Now, ∇C_1 at this point again is going to be radially outwards.

So, what is this telling me about where f can lie? The perpendicular line to this. So, let us take another color here. May not be very visible. So, switch to orange.

Is it visible? Yeah. So, let us say for S. That means S has to be, of course S, if I am saying S is along the tangent, what further condition you think will apply on S? Should be very very small, infinitely small because if I make it big I will leave C_1 which I do not want to do. But of course, all of this is in the limit as for a small quantity. So I can approximately say that this is going to be fine. Is this enough or something else also needs to be looked at? So far I have commented on feasibility, but why am I doing this whole exercise? To go to a better position, right? So that has not yet been brought into the picture.

So to bring this into the picture, a better point means what? $f(x+s) < f(x)$. Only then is it an improvement, right? So this is what is required. Okay. What trick do you think I can apply to simplify this further? Same trick which is Taylor's theorem, right.

So, Taylor's theorem will give me $f(x+s) = f(x) + \nabla f^T s$, okay. What does this imply? Obviously, $\nabla f^T S$ should be less than 0 so that $f(x+S)$ gets reduced with respect to $f(x)$. So, $\nabla f(x)^T S < 0$. Does this remind you of something that you have already studied? It is another way of saying S is a descent direction which makes sense. But that is the algebraic way of looking at it.

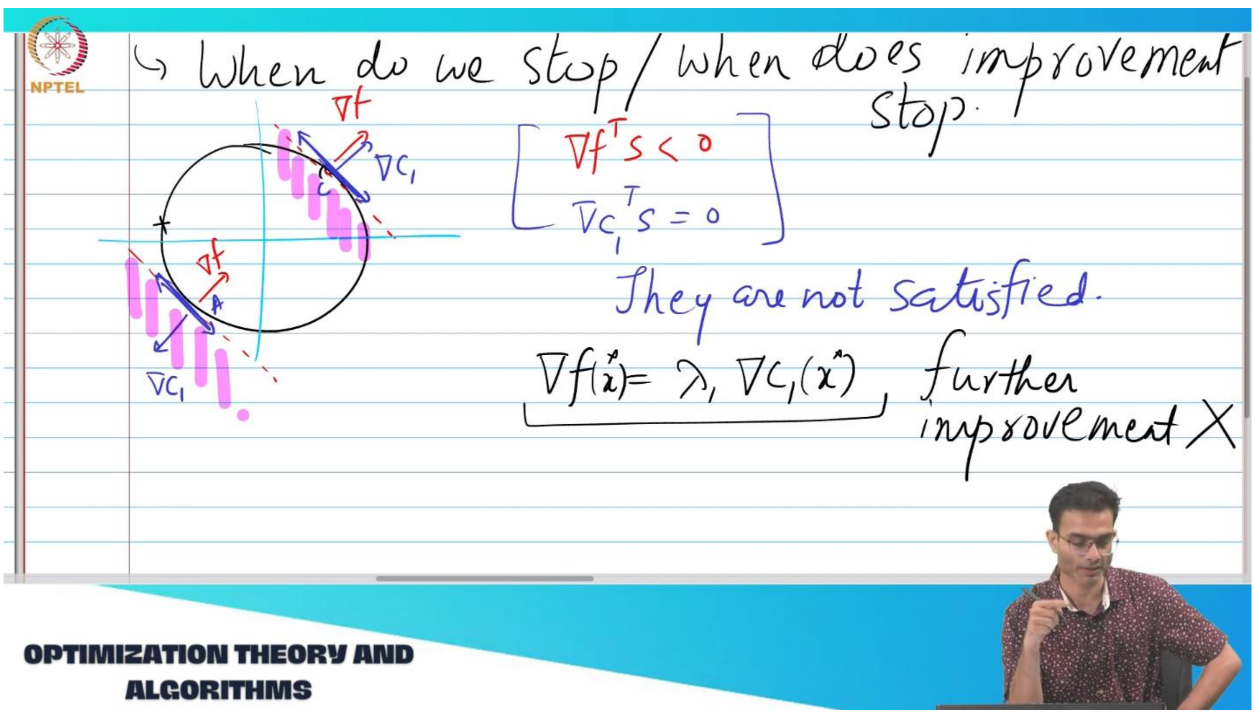


Let us go back to our diagram over here. Which way was my ∇f? ∇f was always in the same direction (1, 1), right? If I use the red color for ∇f and here comes the interesting part, right? So, here is my ∇f. And what is my condition? What? $\nabla f^T S$ should be less than 0. So, how do I quantify this? The geometry of $\nabla f^T S < 0$ is basically a half-plane, a half-space. So, if I were to draw something like this, okay.

Let me try the highlighter. Basically, this region is where S should be. If I take a candidate vector that looks like this, let us take black again, oops. Let us call this S, okay. Does this S satisfy this condition of decrease of the function? It does, because any S that I take in this pink shaded region will satisfy $\nabla f^T S < 0$. So this is a suitable candidate, right? But does this, this S which I have drawn, does it satisfy the feasibility condition? Yes or no? No, because $\nabla C_1^T S \neq 0$.

But it gives us a clue that of the two possible directions that I had for S, which one should I choose? The one in the pink region, right. So, you can see geometrically it is very easy to see. So, it is saying that this direction over here, this is where you should go because it is satisfying both the conditions. In fact, there is a certain nice similarity with gradient descent. I am going exactly in the minus ∇f direction so far, okay.

The two conditions that I need to keep in mind simultaneously are $\nabla C_1^T S = 0$ and $\nabla f^T S < 0$, okay. Remember I want this to be strictly less than 0 because if it is not strictly less, then the function is not improving. Then I am just literally going around in circles. So, this is strictly less than 0.



Is this clear what we did over here? We are just looking at the various things that have to be satisfied. There are only two things that I need to worry about.

One is the constraint being satisfied. That means I stay on the feasible set. That gave me that I should move orthogonal. So, look at the geometry. I should move orthogonal to the gradient of the constraint. That makes sense, right? Because if I move orthogonal to the gradient, supposing I am walking on a hill and I move orthogonal to the gradient of the hill, doesn't my height remain the same? So, same thing, if I am moving orthogonal to the gradient of the constraint, that means the constraint value is remaining the same, which is what I am doing. So I have got my $\nabla C_1^T S = 0$ in plain words and I want the function to improve, I already know what that means, it should be a descent direction. So, $\nabla f^T S < 0$. Putting it together, I have got, if I were at this point, I have a legitimate non-zero direction in which if I go, my function will decrease. So, I can keep doing this in small steps, right.

We are currently not talking about an algorithm. We are only talking about what is possible. We will subsequently come to your simplest algorithm for constraint optimization but right now we are not getting into it. Now, having understood this, what is the next logical question we should ask? When to stop? I can keep doing this but I should now come up with a signature of when to stop. Another way of saying this is when does improvement stop? That is the question that we are asking, okay.

So, here is our circle which looks more like an ellipse, okay. Let us take this point over here. So, ∇f is here, ∇f is here and my ∇C_1 is here, ∇C_1 is here. So this is ∇C_1, ∇C_1, ∇f, ∇f. What were the two conditions that we had? $\nabla C_1^T S$ should be equal to 0, right.

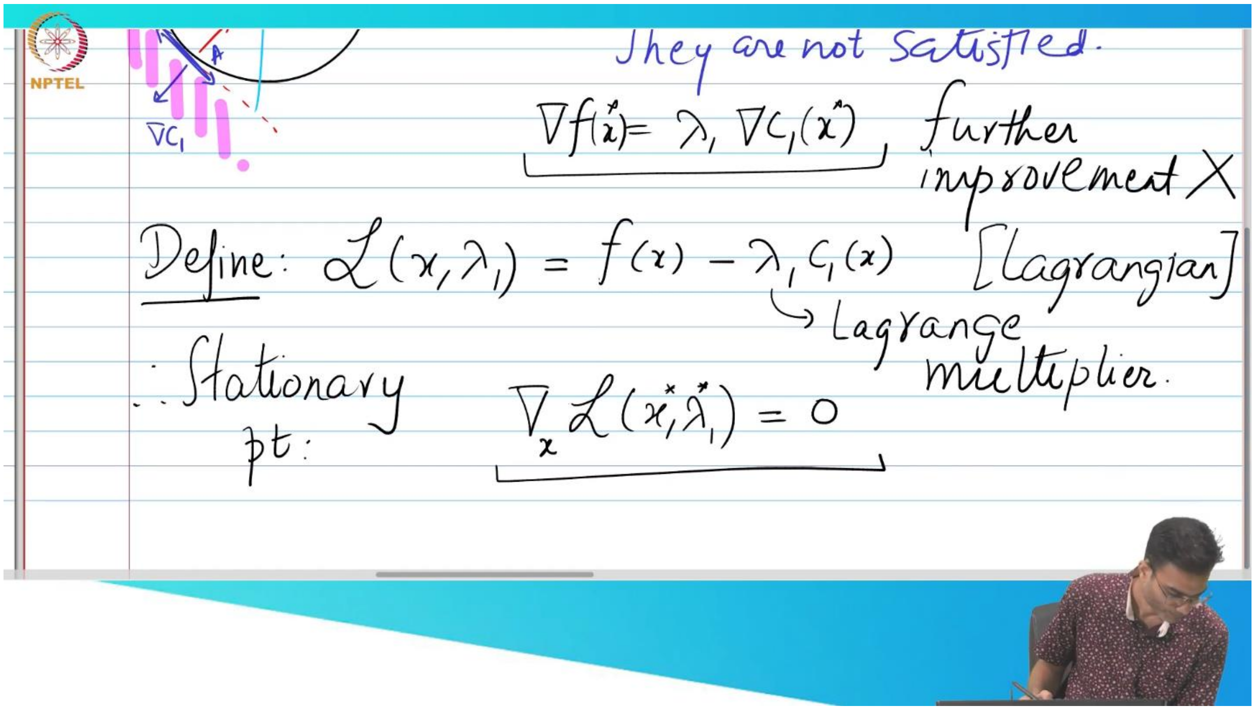
That means I am, if I were to draw a plane like this, I am in which part? Which part? Bottom left, top right. So, let us write $\nabla f^T S < 0$, which part am I? Bottom, right. So, if I were to use the highlighter here, this is where I need to be. For the lower point, $\nabla f^T S$, again it will be the bottom.

So, this is where I need to be. And the second condition was $\nabla C_1^T S = 0$, which tells me I have to be on this line. Is it possible if I am at either of these two points, these points were points A and C, right? This was my point A, this was my point C. Are these two conditions being satisfied at either A or C? I hear a no, anyone for a yes? Take a guess, are they being satisfied? No, because it is the strict inequality which is stopping us, right. So, they are not satisfied. And that is because of the strictness of the inequality, right.

So, we have basically got a signature for when to stop, right. At the previous point, you saw that the location which I had drawn it was possible to find a feasible direction. Now I cannot do that anymore. So graphically what can you say is the relation between ∇f and ∇C at either of these two points? They again seem to be either parallel or anti-parallel, right. I can quantify this as when $\nabla f = \lambda_1 \nabla C_1$.

Then further improvement is not possible. So this is a, you can think of this as a signature for a stationary point, right. Now, you have already a good background in something like steepest descent, right? So, if you were thinking now in terms of an algorithm, hopefully, if you implement it correctly, if you start from some intermediate point, let us say over here by this cross, is it likely for you to end up at the point C or is it more likely you will end up at the point A? You will end up at the point A because from this cross point over here if you follow sufficient, if you follow the first condition, which is $\nabla f^T S < 0$ descent direction, you will never go towards C because that would, in that case, the analog of that is gradient ascent, which we obviously will not be doing. So we are currently not talking about which algorithm we will use but intuitively you know that whatever algorithm, common sense algorithm you use will decrease the function value, and you will end up close to point A.

And point A is actually the true solution. Nevertheless, we can see that mathematically this condition is being satisfied at point C as well. And that is okay because this is the signature for a necessary condition. So that is okay. Now, the way the literature progressed, we want to make a kind of a one-to-one analogy between constrained optimization and unconstrained optimization. In unconstrained optimization, the stationary point is $\nabla f = 0$.



In constraint optimization, is $\nabla f = 0$ the solution? No. There is this lambda business coming over here. So motivated by that to draw a one-to-one analogy people defined a new function. The word you would have heard of many times is called the Lagrangian.

So let us define that. So, it is L as a function of two variables x and λ_1, very simply $L(x, \lambda_1) = f(x) - \lambda_1 C_1(x)$. I have defined it like this, this is called the Lagrangian and this is called the Lagrange multiplier. So now if you got the definition of this Lagrangian correct, now if I ask you what is the signature of a stationary point, what would you say? ∇L with respect to which variable? x, right. So, stationary point. So I will indicate that I am taking derivative with respect to x by putting it as a subscript.

And actually, there will also be an optimum value of λ. So I am saying that this is the signature for a stationary point. So now I can, you can see that it has a kind of a one-to-one symmetry with unconstrained optimization and first order condition. The first order necessary condition for unconstrained optimization was $\nabla f = 0$. The first order necessary condition for constraint optimization is $\nabla L = 0$.

So it is on even footing. Any questions so far? So all we have done is we have looked at feasibility and improvement. These are the two legs on which my algorithm, I mean my intuition is moving forward. These are the two things I want to look at. And that has given me two very simple conditions. What is the constraint over here? I am working with first order Taylor theorem.

So obviously, I have to make small small steps. So that is okay. Okay. So let us move on.